A Bayesian-motivated test for linear model in high-dimensional setting

Rui Wang

Monday 17th December, 2018

1 Introduction

The proposed test is the limit of Bayes factors.

Fixed design

Suppose we would like to test the hypotheses:

$$\mathcal{H}_0 : \mathbf{y} = \mathbf{X}_a\boldsymbol{\beta}_a + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n),$$
$$\mathcal{H}_1 : \mathbf{y} = \mathbf{X}_a\boldsymbol{\beta}_a + \mathbf{X}_b\boldsymbol{\beta}_b + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n).$$

Here $\boldsymbol{\beta}_a$ is q dimensional and $\boldsymbol{\beta}_b$ is p dimensional. We assume that as n tends to infinity, q is fixed while $p/n \to \infty$. This assumption is reasonable. We assume $\mathbf{X}_a$ has full column rank and $\mathbf{X}_b$ has full row rank. In practice, p_0 is often 1 and $\mathbf{X}_a$ is $\mathbf{1}_n$.

As Goeman et al. (2006) pointed out, if $\boldsymbol{\beta}_b \neq 0$ but $\mathbf{X}_b\boldsymbol{\beta}_b = 0$, no test has any power. Goeman et al. (2006) used Bayesian method. Their idea is to choose an 'unbiased' distribution of $\boldsymbol{\beta}_b$. As they noticed, their test has negligible power for many alternatives, and is not unbiased.

The following proposition implies that there is no nontrivial unbiased test.

Proposition 1. *Suppose $\mathbf{y} \sim \mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)$. We test $H_0 : \mu = \mathbf{X}_a\boldsymbol{\beta}_a, \boldsymbol{\beta}_a \in \mathbb{R}^q$ versus $H_1 : \mu \in \mathbb{R}^n$, where $\mathbf{X}_a$ is an $n \times q$ matrix with full column rank, $q < n$. Let $\varphi(\mathbf{y})$ be a test function, that is, a Borel measurable function, $0 \le \phi(\mathbf{y}) \le 1$. If $\int \varphi(\mathbf{y})\mathcal{N}_n(\mathbf{X}_a\boldsymbol{\beta}_a, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) = \alpha$ for $\boldsymbol{\beta}_a \in \mathbb{R}^q$, $\phi > 0$ and $\int \varphi(\mathbf{y})\mathcal{N}_n(\mu, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \ge \alpha$ for $\mu \in \mathbb{R}^n$, $\phi > 0$, then $\varphi(\mathbf{y}) = \alpha$, a.s.*

So we can not find a universally good test. Instead, we would like to find a test with good average behaviour. So Bayesian methods are natural choices in this case.

Bayes hypothesis testing use the Bayes factor.

$$B_{10} = \frac{\int f_1(\mathbf{y}|\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)\pi_1(\boldsymbol{\beta}_b, \boldsymbol{\beta}_a, \phi)d\boldsymbol{\beta}_b d\boldsymbol{\beta}_a d\phi}{\int f_0(\mathbf{y}|\boldsymbol{\beta}_a, \phi)\pi_0(\boldsymbol{\beta}_a, \phi)d\boldsymbol{\beta}_a d\phi}.$$

There have been several extensions of g-priors to $p > n$ case: Maruyama and George (2011), Shang and Clayton (2011).

Under H_0, we impose the reference prior $\pi_0(\boldsymbol{\beta}_a,\phi)=c/\phi$. Note that under H_1, the posterior corresponding to the referece prior is proper if and only if $\mathrm{Rank}(\mathbf{X}_a,\mathbf{X}_b)=q+p$ and $n>q+p$. That is, the minimal training sample size is $q+p+1$. So we cannot impose the reference prior under H_1 provided $q+p\geq n$. We temporarily impose the conditional prior $\boldsymbol{\beta}_b|\boldsymbol{\beta}_a,\phi\sim\mathcal{N}_p(0,\kappa^{-1}\phi^{-1}\mathbf{I}_p)$. There are extansive literature consider the choice of κ. Kass and Wasserman (1995) choose κ such that the amount of information about the parameter equal to the amount of information contained in one observation. Thus, under H_1, we put the prior

$$\pi_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a,\phi)=\mathcal{N}_p\left(0,\frac{1}{\kappa\phi}\mathbf{I}_p\right)(\boldsymbol{\beta}_b),\quad \pi_1(\boldsymbol{\beta}_a,\phi)=\frac{c}{\phi}.$$

$$\begin{aligned}m_0(\mathbf{y};\kappa,\tau)&:=\int f_0^\tau(\mathbf{y}|\boldsymbol{\beta}_a,\phi)\pi_0(\boldsymbol{\beta}_a,\phi)d\boldsymbol{\beta}_a d\phi\\&=\frac{c_0\Gamma\left(\frac{\tau n-q}{2}\right)}{\pi^{\frac{\tau n-q}{2}}\tau^{\frac{\tau n}{2}}|\mathbf{X}_a^\top\mathbf{X}_a|^{\frac{1}{2}}\|(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}\|^{\tau n-q}}.\end{aligned}$$

$$\begin{aligned}m_1(\mathbf{y};\kappa,\tau)&:=\int f_1^\tau(\mathbf{y}|\boldsymbol{\beta}_b,\boldsymbol{\beta}_a,\phi)\pi_1(\boldsymbol{\beta}_b|\boldsymbol{\beta}_a,\phi)\pi_1(\boldsymbol{\beta}_a,\phi)d\boldsymbol{\beta}_a d\boldsymbol{\beta}_b d\phi\\&=\frac{c_1\kappa^{\frac{p}{2}}\Gamma\left(\frac{\tau n-q}{2}\right)}{\pi^{\frac{\tau n-q}{2}}\tau^{\frac{\tau n+p}{2}}|\mathbf{X}_a^\top\mathbf{X}_a|^{\frac{1}{2}}|\mathbf{X}_b^{*\top}\mathbf{X}_b^*+\frac{\kappa}{\tau}\mathbf{I}_p|^{\frac{1}{2}}}\frac{1}{\left[\mathbf{y}^{*\top}\mathbf{y}^*-\mathbf{y}^{*\top}\mathbf{X}_b^*(\mathbf{X}_b^{*\top}\mathbf{X}_b^*+\frac{\kappa}{\tau}\mathbf{I}_p)^{-1}\mathbf{X}_b^{*\top}\mathbf{y}^*\right]^{\frac{\tau n-q}{2}}}.\end{aligned}$$

$$\frac{m_1(\mathbf{y};\kappa,\tau)}{m_0(\mathbf{y};\kappa,\tau)}=\frac{c_1\kappa^{\frac{p}{2}}}{c_0\tau^{\frac{p}{2}}|\mathbf{X}_b^{*\top}\mathbf{X}_b^*+\frac{\kappa}{\tau}\mathbf{I}_p|^{\frac{1}{2}}}\left(\frac{\mathbf{y}^{*\top}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*-\mathbf{y}^{*\top}\mathbf{X}_b^*(\mathbf{X}_b^{*\top}\mathbf{X}_b^*+\frac{\kappa}{\tau}\mathbf{I}_p)^{-1}\mathbf{X}_b^{*\top}\mathbf{y}^*}\right)^{\frac{\tau n-q}{2}}$$

It is straightforward to show that the Bayes factor associated with these priors is

$$\begin{aligned}B_{10}^\kappa=&\frac{\kappa^{p/2}}{|\mathbf{X}_b^\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{X}_b+\kappa\mathbf{I}_p|^{1/2}}\cdot\\&\left(\frac{\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}}{\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}-\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{X}_b\left(\mathbf{X}_b^\top(\mathbf{I}-\mathbf{P}_a)\mathbf{X}_b+\kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}}\right)^{(n-q)/2}.\end{aligned}$$

Thus,

$$\begin{aligned}2\log B_{10}^\kappa=&p\log\kappa-\log|\mathbf{X}_b^\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{X}_b+\kappa\mathbf{I}_p|\\&-(n-q)\log\left(1-\frac{\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{X}_b\left(\mathbf{X}_b^\top(\mathbf{I}-\mathbf{P}_a)\mathbf{X}_b+\kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}}{\mathbf{y}\top(\mathbf{I}_n-\mathbf{P}_a)\mathbf{y}}\right).\end{aligned}$$

Denote by $\mathbf{I}_n-\mathbf{P}_a=\tilde{\mathbf{U}}_a\tilde{\mathbf{U}}_a^\top$ the rank decomposition of $\mathbf{I}_n-\mathbf{P}_a$, where $\tilde{\mathbf{U}}_a$ is a $n\times(n-q)$ column orthogonal matrix. Let $\mathbf{X}_b^*=\tilde{\mathbf{U}}_a^\top\mathbf{X}_b$, $\mathbf{y}^*=\tilde{\mathbf{U}}_a^\top\mathbf{y}$. Let γ_i be the ith largest eigenvalue of $\mathbf{X}_b^*\mathbf{X}_b^{*\top}$, $i=1,\ldots,n-q$. Denote by $\mathbf{X}_b^*=\mathbf{U}_b^*\mathbf{D}_b^*\mathbf{V}_b^{*\top}$ the singular value decomposition of $\mathbf{X}_b^*$, where $\mathbf{U}_b^*$, $\mathbf{V}_b^*$ are $(n-q)\times(n-q)$ and $p\times(n-q)$ column orthogonal matrices, respectively, and $\mathbf{D}_b^*=\mathrm{diag}(\sqrt{\gamma_1},\ldots,\sqrt{\gamma_{n-q}})$. Then

$$
\begin{aligned}
2\log B_{10}^{\kappa} =& p\log\kappa - \sum_{i=1}^{n-q}\log(\gamma_i+\kappa) - (p-(n-q))\log\kappa \\
& - (n-q)\log\left(1 - \frac{\mathbf{y}^{*\top}\mathbf{X}_b^{*}\left(\mathbf{X}_b^{*\top}\mathbf{X}_b^{*}+\kappa\mathbf{I}_p\right)^{-1}\mathbf{X}_b^{*\top}\mathbf{y}^{*}}{\mathbf{y}^{*\top}\mathbf{y}^{*}}\right) \\
=& -\sum_{i=1}^{n-q}\log(\gamma_i+\kappa) + (n-q)\log\left(\frac{\mathbf{y}^{*\top}\mathbf{y}^{*}}{\mathbf{y}^{*\top}\mathbf{U}_b^{*}\left[\frac{1}{\kappa}\left(\mathbf{I}_{n-q} - \mathbf{D}_b^{*}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*}\right)\right]\mathbf{U}_b^{*\top}\mathbf{y}^{*}}\right) \\
=& (n-q)\log\kappa - \sum_{i=1}^{n-q}\log(\gamma_i+\kappa) - (n-q)\log\left(1 - \frac{\mathbf{y}^{*\top}\mathbf{U}_b^{*}\mathbf{D}_b^{*}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*}\mathbf{U}_b^{*\top}\mathbf{y}^{*}}{\mathbf{y}^{*\top}\mathbf{y}^{*}}\right).
\end{aligned}
$$

The main part of $2\log B_{10}^{\kappa}$ is

$$
T_n^{\kappa} = \frac{\mathbf{y}^{*\top}\mathbf{U}_b^{*}\mathbf{D}_b^{*}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*}\mathbf{U}_b^{*\top}\mathbf{y}^{*}}{\mathbf{y}^{*\top}\mathbf{y}^{*}}.
$$

A large value of T_n^{κ} supports the alternative hypothesis. Under the null hypothesis,

$$
\mathrm{E}\,T_n^{\kappa} = \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right).
$$

Under the alternative hypothesis, consider $\boldsymbol{\beta}_b = c\boldsymbol{\beta}_b^{\dagger}$ where $\boldsymbol{\beta}_b^{\dagger} \neq 0$ is a fixed direction and $c > 0$. As $c \to \infty$,

$$
T_n^{\kappa} \to \frac{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^{*}\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^{\dagger}}{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^{*}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^{\dagger}}.
$$

We say T_n^{κ} is consistent along the direction $\boldsymbol{\beta}_b^{\dagger}$ if

$$
\frac{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^{*}\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^{\dagger}}{\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^{*}\mathbf{D}_b^{*2}\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^{\dagger}} > \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right),
$$

or equivalently

$$
\boldsymbol{\beta}_b^{\dagger\top}\mathbf{V}_b^{*}\left[\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2} - \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right)\mathbf{D}_b^{*2}\right]\mathbf{V}_b^{*\top}\boldsymbol{\beta}_b^{\dagger} > 0.
$$

Let k_{κ} be the number of positive eigenvalues of

$$
\mathbf{V}_b^{*}\left[\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2} - \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right)\mathbf{D}_b^{*2}\right]\mathbf{V}_b^{*\top}.
$$

Let $\mathcal{S}_{\kappa}$ be the linear space spanned by the first k_{κ} columns of $\mathbf{V}_b^{*}$. Denote by $\mathcal{S}_{\kappa}^{\perp}$ the orthogonal complement space of $\mathcal{S}_{\kappa}$. We have $\mathbb{R}^p = \mathcal{S}_{\kappa} \oplus \mathcal{S}_{\kappa}^{\perp}$. If $\boldsymbol{\beta}_b^{\dagger} \in \mathcal{S}_{\kappa}$,

$$
\mathbf{V}_b^{*}\left[\mathbf{D}_b^{*2}\left(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q}\right)^{-1}\mathbf{D}_b^{*2} - \frac{1}{n-q}\operatorname{tr}\left(\mathbf{D}_b^{*2}(\mathbf{D}_b^{*2}+\kappa\mathbf{I}_{n-q})^{-1}\right)\mathbf{D}_b^{*2}\right]\mathbf{V}_b^{*\top} > 0.
$$

On the other hand, if $\boldsymbol{\beta}_b^\dagger \in \mathcal{S}_\kappa^\perp$,

$$\mathbf{V}_b^* \left[\mathbf{D}_b^{*2} \left(\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q} \right)^{-1} \mathbf{D}_b^{*2} - \frac{1}{n-q} \operatorname{tr} \left(\mathbf{D}_b^{*2} (\mathbf{D}_b^{*2} + \kappa \mathbf{I}_{n-q})^{-1} \right) \mathbf{D}_b^{*2} \right] \mathbf{V}_b^{*\top} \leq 0.$$

We would like to choose a hyperparameter κ which consists the most consistent directions. To achieve this, we maximize k_κ with respect to κ.

Proposition 2. *For $\kappa_2 > \kappa_1 > 0$, we have $k_{\kappa_1} \geq k_{\kappa_2}$. That is, k_κ ($\kappa > 0$) is decreasing in κ.*

The proposition implies that we should put κ as small as possible. This motivates us to consider $B_{10}^0 = \lim_{\kappa \to 0} B_{10}^\kappa$. It is straightforward to show that

$$2 \log B_{10}^0 = -\sum_{i=1}^{n-q} \log(\gamma_i) + (n-q) \log \left(\frac{\mathbf{y}^{*\top} \mathbf{y}^*}{\mathbf{y}^{*\top} (\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1} \mathbf{y}^*} \right).$$

B_{10}^0 can be regarded as the Bayes factor with respect to noninformative prior.

Define

$$T_n = \frac{\mathbf{y}^{*\top} (\mathbf{X}_b^* \mathbf{X}_b^{*\top})^{-1} \mathbf{y}^*}{\mathbf{y}^{*\top} \mathbf{y}^*}.$$

Then we reject the null hypothesis if T_n is small. It can be seen that under the null hypohtesis,

$$T_n \sim \frac{\sum_{i=1}^{n-q} \gamma_i^{-1} Z_i^2}{\sum_{i=1}^{n-q} Z_i^2},$$

where γ_i is the ith eigenvalue of $\mathbf{X}_b^* \mathbf{X}_b^{*\top}$, $i = 1, \ldots, n-q$, and $Z_1, \ldots, Z_{n-q}$ are iid $\mathcal{N}(0,1)$ random variables.

2 Asymptotic results

Let $\boldsymbol{\varepsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$, where ϵ_i's are iid random variable. Denote $\mu_k = \operatorname{E} \epsilon_1^k$. Then $\mu_1 = 0$, $\mu_2 = \phi^{-1}$.

Assumption 1. *Suppose*

Lemma 1. *If $\phi^2 \mu_4 = o(n-q)$,*

$$\mathbf{y}^{*\top} \mathbf{y}^* = (1 + o_P(1)) \left(\boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b + \phi^{-1}(n-q) \right).$$

Proof.

$$\mathbf{y}^{*\top} \mathbf{y}^* = \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b + 2\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b + \boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \boldsymbol{\varepsilon}.$$

$$\operatorname{E} \left(\mathbf{y}^{*\top} \mathbf{y}^* \right) = \boldsymbol{\beta}_b^\top \mathbf{X}_b^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b + \phi^{-1}(n-q).$$

$$\operatorname{Var} \left(\mathbf{y}^{*\top} \mathbf{y}^* \right) \leq 2 \operatorname{Var} \left(2\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \mathbf{X}_b \boldsymbol{\beta}_b \right) + 2 \operatorname{Var} \left(\boldsymbol{\varepsilon}^\top (\mathbf{I}_n - \mathbf{P}_a) \boldsymbol{\varepsilon} \right)$$

From (i) of (Chen et al., 2010, Proposition A.1),

$$\operatorname{Var}\left(\boldsymbol{\varepsilon}^\top(\mathbf{I}_n - \mathbf{P}_a)\boldsymbol{\varepsilon}\right) = \phi^{-2}\left((\phi^2\mu_4 - 3)\sum_{i=1}^n ((\mathbf{I}_n - \mathbf{P}_a)_{i,i})^2 + 2(n-q)\right) \le \phi^{-2}(2+\phi^2\mu_4)(n-q).$$

Then

$$\operatorname{Var}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right) \le 8\phi^{-1}\boldsymbol{\beta}_b^\top\mathbf{X}_b^\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + 2\phi^{-2}(2+\phi^2\mu_4)(n-q)$$

Thus, if $\phi^2\mu_4 = o(n-q)$, we have

$$\frac{\operatorname{Var}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right)}{\left(\operatorname{E}\left(\mathbf{y}^{*\top}\mathbf{y}^*\right)\right)^2} \to 0,$$

and consequently $\mathbf{y}^{*\top}\mathbf{y}^* = (1 + o_P(1))\operatorname{E}(\mathbf{y}^{*\top}\mathbf{y}^*)$.

□

Note that under the normality, $T_n - \operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})/(n-q)$ has zero mean.

Theorem 1. *Suppose the rows of $\mathbf{X}_b$ are iid random vectors with distribution $\mathcal{N}(0, \sigma_b^2\mathbf{I}_p)$. Suppose $p/(n-q) \to c \in (1, +\infty)$. Then*

$$\left(\boldsymbol{\beta}_b^\top\mathbf{X}_b^\top(\mathbf{I}_n - \mathbf{P}_a)\mathbf{X}_b\boldsymbol{\beta}_b + \phi^{-1}(n-q)\right)\left(\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\right) \rightsquigarrow \mathcal{N}(0,1).$$

Proof. Note that $\mathbf{X}_b^*\mathbf{X}_b^{*\top} \sim \operatorname{Wishart}(p, \sigma_b^2\mathbf{I}_{n-q})$.

$$\frac{\mathbf{y}^{*\top}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\mathbf{y}^*}{\mathbf{y}^{*\top}\mathbf{y}^*} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q} = \frac{\phi\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{y}^*}{\phi\mathbf{y}^{*\top}\mathbf{y}^*}.$$

We have

$$\begin{aligned}
&\phi\mathbf{y}^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{y}^*\\
=&\phi\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\tilde{\mathbf{U}}_a^\top\boldsymbol{\varepsilon}\\
&+ 2\phi\boldsymbol{\varepsilon}^\top\tilde{\mathbf{U}}_a\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{X}_b^*\boldsymbol{\beta}_b\\
&+ \phi\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)\mathbf{X}_b^*\boldsymbol{\beta}_b\\
=:&A_1 + A_2 + A_3.
\end{aligned}$$

We have $\operatorname{E}(A_1|\mathbf{X}_b) = \operatorname{E}(A_2|\mathbf{X}_b) = 0$. It is also straightforward to see that

$$\begin{aligned}
\operatorname{Var}(A_1|\mathbf{X}_b) &= 2\operatorname{tr}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-2}\right) - 2\frac{1}{n-q}\operatorname{tr}^2\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right),\\
\operatorname{Var}(A_2|\mathbf{X}_b) &= 4\boldsymbol{\beta}_b^\top\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\mathbf{X}_b^*\boldsymbol{\beta}_b,\\
\operatorname{Cov}(A_1, A_2|\mathbf{X}_b) &= 0.
\end{aligned}$$

By some theory, we have From (Jiang, 1996, Theorem 5.1),

$$\frac{A_1 + A_2}{\sqrt{\operatorname{Var}(A_1|\mathbf{X}_b) + \operatorname{Var}(A_2|\mathbf{X}_b)}} \rightsquigarrow \mathcal{N}(0,1).$$

From lemma 3,

$$\operatorname{Var}(A_1|\mathbf{X}_b) = (1+o_P(1))2\sigma_b^{-4}p^{-2}(n-q)\operatorname{Var}(\xi^{-2}).$$

Let $\mathbf{O}$ be a $p \times p$ random matrix with Haar distribution which is independent of $\mathbf{X}_b$. The rotation invariance of normal distribution implies that $\mathbf{X}_b\mathbf{O}$ has the same distribution as $\mathbf{X}_b$ and is independent of $\mathbf{O}$. Then

$$\begin{aligned}
\operatorname{Var}(A_2|\mathbf{X}_b) &= 4\boldsymbol{\beta}_b^\top \mathbf{O}\mathbf{O}^\top \mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2 \mathbf{X}_b^*\mathbf{O}\mathbf{O}^\top\boldsymbol{\beta}_b \\
&\overset{d}{=} 4\boldsymbol{\beta}_b^\top \mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2 \mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b.
\end{aligned}$$

Note that $\mathbf{O}^\top\boldsymbol{\beta}_b/\|\mathbf{O}^\top\boldsymbol{\beta}_b\|$ is uniformly distributed on the unit sphere S^{p-1}. From Lemma 2,

$$\begin{aligned}
&\operatorname{E}\left(4\boldsymbol{\beta}_b^\top \mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2 \mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b \middle| \mathbf{X}_b\right) \\
=&4p^{-1}\|\boldsymbol{\beta}_b\|^2 \operatorname{tr}\left(\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2 \mathbf{X}_b^*\right) \\
=&4p^{-1}\|\boldsymbol{\beta}_b\|^2 \left(\frac{1}{(n-q)^2}\operatorname{tr}^2\left[(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right]\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top}) - \operatorname{tr}\left[(\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}\right]\right).
\end{aligned}$$

Then Lemma 3 implies that

$$\begin{aligned}
&\operatorname{E}\left(4\boldsymbol{\beta}_b^\top \mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2 \mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b \middle| \mathbf{X}_b\right) \\
=&(1+o_P(1))4\|\boldsymbol{\beta}_b\|^2\sigma_b^{-2}p^{-2}(n-q)\left(\operatorname{E}(\xi)\left(\operatorname{E}(\xi^{-1})\right)^2 - \operatorname{E}(\xi^{-1})\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\operatorname{Var}\left(4\boldsymbol{\beta}_b^\top \mathbf{O}\mathbf{X}_b^{*\top}\left((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1} - \frac{\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2 \mathbf{X}_b^*\mathbf{O}^\top\boldsymbol{\beta}_b \middle| \mathbf{X}_b\right) \\
\leq&\frac{32}{p^2}\|\boldsymbol{\beta}_b\|^4\Bigg(\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-2}) + \frac{1}{(n-q)^2}\operatorname{tr}^4((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})\operatorname{tr}((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^2) \\
&+ 2\operatorname{tr}^2((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1}) - \frac{4}{(n-q)^2}\operatorname{tr}^3((\mathbf{X}_b^*\mathbf{X}_b^{*\top})^{-1})\operatorname{tr}(\mathbf{X}_b^*\mathbf{X}_b^{*\top})\Bigg) \\
=&(1+o_P(1))32\|\boldsymbol{\beta}_b\|^4\sigma_b^{-4}p^{-4} \\
&\left(\left((n-q)\operatorname{E}(\xi^{-2}) + (n-q)^3(\operatorname{E}(\xi^{-1}))^4\operatorname{E}(\xi^2) + 2(n-q)^2(\operatorname{E}(\xi^{-1}))^2 - 4(n-q)^2(\operatorname{E}(\xi^{-1}))^3\operatorname{E}(\xi)\right)\right)
\end{aligned}$$

$\square$

Lemma 2. *Let $\mathbf{A}$ be an $p \times p$ symmetric matrix. Let Z be a p dimensional random vector with uniform distribution on the unit sphere S^{p-1}. Then*

$$\mathrm{E}(Z^\top \mathbf{A} Z) = \frac{1}{p}\operatorname{tr}(\mathbf{A}), \quad \operatorname{Var}(Z^\top \mathbf{A} Z) = \frac{2}{p(p+2)}\left(\operatorname{tr}(\mathbf{A}^2) - \frac{1}{p}\operatorname{tr}^2(\mathbf{A})\right) \le \frac{2}{p^2}\operatorname{tr}(\mathbf{A}^2).$$

Proof. The result follows from direct calculation and the fact that for nonnegative integers $k_1, \ldots, k_p$,

$$\mathrm{E}\prod_{i=1}^p z_i^{2k_i} = \frac{\Gamma(p/2)\prod_{i=1}^p \Gamma(k_i + 1/2)}{\pi^{p/2}\Gamma(\sum_{i=1}^p k_i + p/2)},$$

where z_i is the ith coordinate of Z. □

The following lemma is a direct consequence of MP law and Bai Yin law.

Lemma 3. *Under the assumptions of Theorem 1, for every $r \in \mathbb{R}$,*

$$\frac{1}{\sigma_b^{2r} p^r (n-q)}\operatorname{tr}((\mathbf{X}_b^* \mathbf{X}_b^{*\top})^r) \xrightarrow{a.s.} \mathrm{E}\,\xi^r,$$

where ξ is a random variable with density function

$$p_c(x) = \mathbf{1}_{\left[(1-c^{-1/2})^2, (1+c^{-1/2})^2\right]}(x)\frac{c}{2\pi x}\sqrt{4/c - (x - (1/c+1))^2}.$$

As in Vershynin (2018), sub-gaussian norm of a sub-gaussian random variable is defined as

$$\|X\|_{\psi_2} = \inf\{t > 0 : \mathrm{E}\exp(X^2/t^2) \le 2\}.$$

A random vector $Z \in \mathbb{R}^p$ is called sub-gaussian if $z \top Z$ are sub-gaussian random variables for all $z \in \mathbb{R}^p$. The sub-gaussian norm of Z is defined as

$$\|Z\|_{\psi_2} = \sup_{z \in S^{p-1}} \|z^\top Z\|_{\psi_2},$$

where S^{p-1} is the unit sphere in $\mathbb{R}^p$.

Suppose $\mathbf{X}_b = \mathbf{Z}_b \Gamma + \mathbf{1}_n \mu_b^\top$, where the rows of $\mathbf{Z}_b$ are iid sub-gaussion random vectors with identity covariance matrix.

The following lemma is a simple extension of Theorem 4.6.1 of Vershynin (2018).

Lemma 4. *Let $\mathbf{Z}$ be an $N \times n$ random matrix whose columns Z_i are independent sub-gaussian random vectors with $\mathrm{E}(Z_i) = 0$, $\operatorname{Var}(Z_i) = \mathbf{I}_n$. Suppose $K := \max_i \|Z_i\|_{\psi_2}$ is uniformly bounded. Write $Z_i = (z_{i1}, \ldots, z_{iN})^\top$. Assume that $\mathrm{E}(z_{i\ell}^4) = 3 + \Delta < \infty$ and for any intergers $\ell_v \ge 0$ with $\sum_{v=1}^s \ell_v \le 4$,*

$$\mathrm{E}(Z_{ij_1}^{\ell_1} Z_{ij_2}^{\ell_2} \cdots Z_{ij_s}^{\ell_s}) = \mathrm{E}(Z_{ij_1}^{\ell_1})\,\mathrm{E}(Z_{ij_2}^{\ell_2}) \cdots \mathrm{E}(Z_{ij_s}^{\ell_s})$$

Let $\mathbf{W}$ be a nonrandom $N \times N$ symmetric matrix. Then

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \operatorname{tr}(\mathbf{W})\mathbf{I}_n\| = O_P(\sqrt{n}\|\mathbf{W}\|_F + n\|\mathbf{W}\|).$$

TO BE DONE:

$$\|U\mathbf{Z}^\top \mathbf{W} \mathbf{Z} U^\top - \operatorname{tr}(\mathbf{W})\mathbf{I}_n\|$$

Next we verify the following. Let $\mathbf{\Sigma} = \Gamma^\top \Gamma$.

Proof. Let

$$\mathbf{B} = \mathbf{X}_b^* \mathbf{X}_b^{*\top} = \tilde{\mathbf{U}}_a^\top \mathbf{X}_b \mathbf{X}_b^\top \tilde{\mathbf{U}}_a = \tilde{\mathbf{U}}_a^\top \mathbf{Z}_b \Gamma \Gamma^\top \mathbf{Z}_b^\top \tilde{\mathbf{U}}_a$$

Note that Lemma 4 implies that

$$\|\mathbf{B} - \mathrm{tr}(\mathbf{\Sigma}) \mathbf{I}_{n-q}\| = O_P(\sqrt{n}\|\mathbf{\Sigma}\|_F + n\|\mathbf{\Sigma}\|).$$

That is, uniformly for $i = 1, \ldots, n-q$,

$$\frac{\lambda_i(\mathbf{B})}{\mathrm{tr}(\mathbf{\Sigma})} = 1 + O_P\left(\frac{\sqrt{n}\|\mathbf{\Sigma}\|_F}{\mathrm{tr}(\mathbf{\Sigma})} + \frac{n\|\mathbf{\Sigma}\|}{\mathrm{tr}(\mathbf{\Sigma})} \right)$$

Define

$$\delta_i = \frac{\lambda_i(\mathbf{B})}{\mathrm{tr}(\mathbf{\Sigma})} - 1$$

$$\eta = \frac{\sqrt{n}\|\mathbf{\Sigma}\|_F}{\mathrm{tr}(\mathbf{\Sigma})} + \frac{n\|\mathbf{\Sigma}\|}{\mathrm{tr}(\mathbf{\Sigma})}$$

We assume $\eta \to 0$.

Thus, we need to verify

$$\tilde{\mathbf{U}}_a \mathbf{B}^{-1} \tilde{\mathbf{U}}_a^\top - \frac{\mathrm{tr}(\mathbf{B}^{-1})}{n-q} \tilde{\mathbf{U}}_a \tilde{\mathbf{U}}_a^\top$$

satisfies that

$$\left\| \mathbf{B}^{-1} - \frac{\mathrm{tr}(\mathbf{B}^{-1})}{n-q} \mathbf{I}_{n-q} \right\|^2 \Big/ \mathrm{tr}\left(\mathbf{B}^{-1} - \frac{\mathrm{tr}(\mathbf{B}^{-1})}{n-q} \mathbf{I}_{n-q} \right)^2 \to 0. \tag{1}$$

Note that

$$\mathrm{tr}\left(\mathbf{B}^{-1} - \frac{\mathrm{tr}(\mathbf{B}^{-1})}{n-q} \mathbf{I}_{n-q} \right)^2 = \sum_{i=1}^{n-q} \frac{1}{\lambda_i^2(\mathbf{B})} - \frac{1}{n-q} \left(\sum_{i=1}^{n-q} \frac{1}{\lambda_i(\mathbf{B})} \right)^2.$$

By Taylor's theorem, uniformly for $i = 1, \ldots, n$, we have

$$\begin{aligned}
\frac{1}{\lambda_i(\mathbf{B})} &= \frac{1}{\mathrm{tr}(\mathbf{\Sigma})} \frac{1}{1+\delta_i} = \frac{1}{\mathrm{tr}(\mathbf{\Sigma})} \left(1 - \delta_i + \delta_i^2 + O_P(\eta^3) \right), \\
\frac{1}{\lambda_i^2(\mathbf{B})} &= \frac{1}{\mathrm{tr}^2(\mathbf{\Sigma})} \frac{1}{(1+\delta_i)^2} = \frac{1}{\mathrm{tr}^2(\mathbf{\Sigma})} \left(1 - 2\delta_i + 3\delta_i^2 + O_P(\eta^3) \right).
\end{aligned}$$

Thus,

$$
\begin{aligned}
&\operatorname{tr}\left(\mathbf{B}^{-1}-\frac{\operatorname{tr}(\mathbf{B}^{-1})}{n-q}\mathbf{I}_{n-q}\right)^2\\
=&\sum_{i=1}^{n-q}\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(1-2\delta_i+3\delta_i^2+O_P(\eta^3)\right)-\frac{1}{n-q}\left(\sum_{i=1}^{n-q}\frac{1}{\operatorname{tr}(\mathbf{\Sigma})}\left(1-\delta_i+\delta_i^2+O_P(\eta^3)\right)\right)^2\\
=&\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(n-q-2\sum_{i=1}^{n-q}\delta_i+3\sum_{i=1}^{n-q}\delta_i^2+O_P(n\eta^3)-\frac{1}{n-q}\left(n-q-\sum_{i=1}^{n-q}\delta_i+\sum_{i=1}^{n-q}\delta_i^2+O_P(n\eta^3)\right)^2\right)\\
=&\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(n-q-2\sum_{i=1}^{n-q}\delta_i+3\sum_{i=1}^{n-q}\delta_i^2+O_P(n\eta^3)-(n-q)\left(1-\frac{1}{n-q}\sum_{i=1}^{n-q}\delta_i+\frac{1}{n-q}\sum_{i=1}^{n-q}\delta_i^2+O_P(\eta^3)\right)^2\right)\\
=&\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\Bigg(n-q-2\sum_{i=1}^{n-q}\delta_i+3\sum_{i=1}^{n-q}\delta_i^2+O_P(n\eta^3)\\
&-(n-q)\left(1+\left(\frac{1}{n-q}\sum_{i=1}^{n-q}\delta_i\right)^2-\frac{2}{n-q}\sum_{i=1}^{n-q}\delta_i+\frac{2}{n-q}\sum_{i=1}^{n-q}\delta_i^2+O_P(\eta^3)\right)\Bigg)\\
=&\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(\sum_{i=1}^{n-q}\delta_i^2-\frac{1}{n-q}\left(\sum_{i=1}^{n-q}\delta_i\right)^2+O_P(n\eta^3)\right)\\
=&\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(\operatorname{tr}\left(\frac{1}{\operatorname{tr}(\mathbf{\Sigma})}\mathbf{B}-\mathbf{I}_{n-q}\right)^2-\frac{1}{n-q}\left(\frac{\operatorname{tr}(\mathbf{B})}{\operatorname{tr}(\mathbf{\Sigma})}-(n-q)\right)^2+O_P(n\eta^3)\right)\\
=&\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(\frac{1}{\operatorname{tr}^2(\mathbf{\Sigma})}\left(\operatorname{tr}(\mathbf{B}^2)-\frac{1}{n-q}\operatorname{tr}^2(\mathbf{B})\right)+O_P(n\eta^3)\right).
\end{aligned}
$$

□

Lemma 5. *Let $\mathbf{Z}$ be an $n\times m$ random matrix whose rows Z_i are independent random vectors with $\mathrm{E}(Z_i)=0$, $\mathrm{Var}(Z_i)=\mathbf{I}_m$. Write $Z_i=(z_{i1},\ldots,z_{im})^\top$. Assume that $\mathrm{E}(z_{i\ell}^4)=3+\Delta<\infty$ and for any intergers $\ell_v\geq 0$ with $\sum_{v=1}^{s}\ell_v\leq 8$. Let $\mathbf{B}$ be a $m\times m$ symmetric matrix. Let $\mathbf{Q}$ be a $n\times n$ projection maitrx with rank r. Then*

$$
\operatorname{tr}(\mathbf{Q}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})^2-\frac{1}{r}\left(\operatorname{tr}(\mathbf{Q}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})\right)^2
$$

Proof. Let $\mathbf{Q}=\mathbf{U}\mathbf{U}^\top$, where $\mathbf{U}$ is a $n\times r$ column orthogonal matrix. Then

$$
\begin{aligned}
&\operatorname{tr}(\mathbf{Q}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})^2-\frac{1}{r}\left(\operatorname{tr}(\mathbf{Q}\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})\right)^2\\
=&\operatorname{tr}(\mathbf{U}^\top\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{U})^2-\frac{1}{r}\left(\operatorname{tr}(\mathbf{U}^\top\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{U})\right)^2=
\end{aligned}
$$

$$
\operatorname{tr}(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})=\sum_{i=1}^{n}\sum_{j=1}^{n}q_{i,j}Z_i^\top\mathbf{B}Z_j
$$

$$\operatorname{tr}^2(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q}) = \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n q_{i,j}q_{k,l}Z_i^\top\mathbf{B}Z_jZ_k^\top\mathbf{B}Z_l$$

And

$$\operatorname{tr}(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})^2 = \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n q_{j,k}q_{l,i}Z_i^\top\mathbf{B}Z_jZ_k^\top\mathbf{B}Z_l$$

Thus,

$$\operatorname{tr}(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q})^2 - r^{-1}\operatorname{tr}^2(\mathbf{Z}\mathbf{B}\mathbf{Z}^\top\mathbf{Q}) = \sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n\sum_{l=1}^n (q_{i,l}q_{j,k} - r^{-1}q_{i,j}q_{k,l})Z_i^\top\mathbf{B}Z_jZ_k^\top\mathbf{B}Z_l$$

Thus,

$$e(i,j,k,l) := q_{i,l}q_{j,k} - r^{-1}q_{i,j}q_{k,l}$$

satisfies,

$$e(i_1,i_2,i_3,i_4) = e(i_2,i_1,i_4,i_3) = e(i_3,i_4,i_1,i_2) = e(i_4,i_3,i_2,i_1)$$

□

Lemma 6. *Let $e(i,j;k,l)$ satisfy:*

$$e(i_1,i_2,i_3,i_4) = e(i_2,i_1,i_4,i_3) = e(i_3,i_4,i_1,i_2) = e(i_4,i_3,i_2,i_1)$$

Then

$$\sum_{i,j,k,l=1}^n e(i,j,k,l)Z_i^\top\mathbf{B}Z_jZ_k^\top\mathbf{B}Z_l$$

Proof. Let $f(i,j,k,l) = e(i,j,k,l)Z_i^\top\mathbf{B}Z_jZ_k^\top\mathbf{B}Z_l$. Then f also satisfies

$$f(i_1,i_2,i_3,i_4) = f(i_2,i_1,i_4,i_3) = f(i_3,i_4,i_1,i_2) = f(i_4,i_3,i_2,i_1)$$

$$\begin{aligned}
&\sum_{i,j,k,l=1}^n f(i,j,k,l)\\
&= \sum_{i=j=1}^n\sum_{k=l=1}^n + \sum_{i=j=1}^n\sum_{k,l}^* + \sum_{i,j}^*\sum_{k=l=1}^n + \sum_{i,j}^*\sum_{k,l}^*\\
&= \sum_{i=1}^n\sum_{j=1}^n f(i,i,j,j) + \sum_{i=1}^n\sum_{j,k}^* f(i,i,j,k) + \sum_{i=1}^n\sum_{j,k}^* f(j,k,i,i) + \sum_{i,j}^*\sum_{k,l}^* f(i,j,k,l)\\
&= \sum_{i=1}^n\sum_{j=1}^n f(i,i,j,j) + 2\sum_{i=1}^n\sum_{j,k}^* f(i,i,j,k) + \sum_{i,j}^*\sum_{k,l}^* f(i,j,k,l)
\end{aligned}$$

For the first term,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} f(i,i,j,j) = \sum_{i=1}^{n} f(i,i,i,i) + \sum_{i,j}^{*} f(i,i,j,j).$$

For the second term,

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{j,k}^{*} f(i,i,j,k) \\
=&\sum_{i=1}^{n}\sum_{j\neq i}^{n} f(i,i,i,j) + \sum_{i=1}^{n}\sum_{j\neq i} f(i,i,j,i) + \sum_{i,j,k}^{*} f(i,i,j,k) \\
=&2\sum_{i,j}^{*} f(i,i,i,j) + \sum_{i,j,k}^{*} f(i,i,j,k)
\end{aligned}$$

For the third term,

$$\begin{aligned}
&\sum_{i,j}^{*}\sum_{k,l}^{*} f(i,j,k,l) \\
=&\sum_{i,j}^{*}\sum_{k=i,l=j} + \sum_{i,j}^{*}\sum_{k=j,l=i} + \sum_{i,j}^{*}\sum_{k=i,l\neq\{i,j\}} + \sum_{i,j}^{*}\sum_{k=j,l\neq\{i,j\}} + \sum_{i,j}^{*}\sum_{l=i,k\neq\{i,j\}} + \sum_{i,j}^{*}\sum_{l=j,k\neq\{i,j\}} + \sum_{i,j,k,l}^{*} \\
=&\sum_{i,j}^{*}(f(i,j,i,j) + f(i,j,j,i)) + 2\sum_{i,j,k}^{*}(f(i,j,i,k) + f(i,j,k,i)) + \sum_{i,j,k,l}^{*} f(i,j,k,l)
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{i,j,k,l=1}^{n} f(i,j,k,l) =& \sum_{i=1}^{n} f(i,i,i,i) + \sum_{i,j}^{*} f(i,i,j,j) \\
&+ 4\sum_{i,j}^{*} f(i,i,i,j) + 2\sum_{i,j,k}^{*} f(i,i,j,k) \\
&\sum_{i,j}^{*}(f(i,j,i,j) + f(i,j,j,i)) + 2\sum_{i,j,k}^{*}(f(i,j,i,k) + f(i,j,k,i)) + \sum_{i,j,k,l}^{*} f(i,j,k,l) \\
=:& \sum_{i=1}^{7} A_i
\end{aligned}$$

Note that for distinct i,j,k,l,

$$\begin{aligned}
&\mathrm{E}\, f(i,i,i,i) = e(i,i,i,i)\,\mathrm{E}(Z_i^\top \mathbf{B} Z_i)^2 = e(i,i,i,i)\left(\operatorname{tr}^2(\mathbf{B}) + 2\operatorname{tr}(\mathbf{B}^2) + \Delta\operatorname{tr}(\mathbf{B}\circ\mathbf{B})\right) \\
&\mathrm{E}\, f(i,i,j,j) = e(i,i,j,j)\,\mathrm{E}(Z_i^\top \mathbf{B} Z_i)(Z_i^\top \mathbf{B} Z_i) = e(i,i,j,j)\operatorname{tr}^2(\mathbf{B}) \\
&\mathrm{E}\, f(i,i,i,j) = 0 \\
&\mathrm{E}\,(f(i,j,i,j) + f(i,j,j,i)) = (e(i,j,i,j) + e(i,j,j,i))\,\mathrm{E}(Z_i^\top \mathbf{B} Z_j)^2 = (e(i,j,i,j) + e(i,j,j,i))\,\mathrm{E}(\mathbf{B}^2) \\
&\mathrm{E}\,(f(i,j,i,k) + f(i,j,k,i)) = 0 \\
&\mathrm{E}\, f(i,j,k,l) = 0
\end{aligned}$$

Let $h(i,j,k,l) = f(i,j,k,l) - \mathrm{E}\, f(i,j,k,l)$. Then

$$\begin{aligned}\mathrm{Var}\left(\sum_{i,j,k,l=1}^{n} e(i,j,k,l) Z_i^\top \mathbf{B} Z_j Z_k^\top \mathbf{B} Z_l\right) &= \mathrm{E}\left(\sum_{i,j,k,l=1}^{n} h(i,j,k,l)\right)^2 \\ &= \sum_{i=1}^{7} \mathrm{E}(A_i - \mathrm{E}\, A_i)^2 + 2 \sum_{1\le i<j\le 7} \mathrm{E}(A_i - \mathrm{E}\, A_i)(A_i - \mathrm{E}\, A_i)\end{aligned}$$

We have

$$\mathrm{E}(A_1 - \mathrm{E}\, A_1)^2 = \sum_{i=1}^{n} e(i,i,i,i)^2\, \mathrm{E}\left((Z_1^\top \mathbf{B} Z_1)^2 - \mathrm{E}(Z_1^\top \mathbf{B} Z_1)^2\right)^2$$

□

Lemma 7.

$$\mathrm{E}\left((Z^\top \mathbf{B} Z)^2 - \mathrm{E}(Z^\top \mathbf{B} Z)^2\right)^2 =$$

Proof.

$$\begin{aligned}(Z^\top \mathbf{B} Z)^2 &= \left(\sum_{i=1}^{m} b_{i,i} z_i^2 + \sum_{i,j}^{*} b_{i,j} z_i z_j\right)^2 \\ &= \left(\sum_{i=1}^{m} b_{i,i} z_i^2\right)^2 + \left(\sum_{i,j}^{*} b_{i,j} z_i z_j\right)^2 + 2\left(\sum_{i=1}^{m} b_{i,i} z_i^2\right)\left(\sum_{i,j}^{*} b_{i,j} z_i z_j\right) \\ &= \sum_{i=1}^{m} b_{i,i}^2 z_i^4 + \sum_{i,j}^{*} b_{i,i} b_{j,j} z_i^2 z_j^2 \\ &\quad + \sum_{i,j}^{*} \sum_{k,l}^{*} b_{i,j} b_{k,l} z_i z_j z_k z_l + 2 \sum_{k=1}^{m} \sum_{i,j}^{*} b_{i,j} b_{k,k} z_i z_j z_k^2\end{aligned}$$

Note that

$$\begin{aligned}&\sum_{i,j}^{*} \sum_{k,l}^{*} b_{i,j} b_{k,l} z_i z_j z_k z_l \\ =&\sum_{i,j}^{*} \sum_{k=i,l=j}^{*} + \sum_{i,j}^{*} \sum_{k=i,l\neq j}^{*} + \sum_{i,j}^{*} \sum_{k=j,l=i}^{*} + \sum_{i,j}^{*} \sum_{k=j,l\neq i}^{*} + \sum_{i,j}^{*} \sum_{k\notin\{i,j\},l=i}^{*} + \sum_{i,j}^{*} \sum_{k\notin\{i,j\},l=j}^{*} + \sum_{i,j}^{*} \sum_{k\notin\{i,j\},l\notin\{i,j\}}^{*} \\ =&2\sum_{i,j}^{*} b_{i,j}^2 z_i^2 z_j^2 + 4\sum_{i,j,k}^{*} b_{i,j} b_{i,k} z_i^2 z_j z_k + \sum_{i,j,k,l}^{*} b_{i,j} b_{k,l} z_i z_j z_k z_l\end{aligned}$$

On the other hand

$$\begin{aligned}&\sum_{k=1}^{m} \sum_{i,j}^{*} b_{i,j} b_{k,k} z_i z_j z_k^2 = \sum_{i,j}^{*} \sum_{k=1}^{m} b_{i,j} b_{k,k} z_i z_j z_k^2 \\ =&\sum_{i,j}^{*} b_{i,j} b_{i,i} z_i^3 z_j + \sum_{i,j}^{*} b_{i,j} b_{j,j} z_i z_j^3 + \sum_{i,j,k}^{*} b_{i,j} b_{k,k} z_i z_j z_k^2 = \sum_{i,j}^{*} b_{i,i} b_{i,j} z_i^3 z_j + \sum_{i,j,k}^{*} b_{i,j} b_{k,k} z_i z_j z_k^2.\end{aligned}$$

Thus,

$$
\begin{aligned}
(Z^\top \mathbf{B} Z)^2 &= \sum_{i=1}^m b_{i,i}^2 z_i^4 + \sum_{i,j}^* b_{i,i} b_{j,j} z_i^2 z_j^2 + \sum_{i,j}^* \sum_{k,l}^* b_{i,j} b_{k,l} z_i z_j z_k z_l + 2 \sum_{k=1}^m \sum_{i,j}^* b_{i,j} b_{k,k} z_i z_j z_k^2 \\
&= \sum_{i=1}^m b_{i,i}^2 z_i^4 + \sum_{i,j}^* b_{i,i} b_{j,j} z_i^2 z_j^2 + 2 \sum_{i,j}^* b_{i,j}^2 z_i^2 z_j^2 + 4 \sum_{i,j,k}^* b_{i,j} b_{i,k} z_i^2 z_j z_k + \sum_{i,j,k,l}^* b_{i,j} b_{k,l} z_i z_j z_k z_l \\
&\quad + 2 \sum_{i,j}^* b_{i,i} b_{i,j} z_i^3 z_j + 2 \sum_{i,j,k}^* b_{i,j} b_{k,k} z_i z_j z_k^2. \\
&= \sum_{i=1}^m b_{i,i}^2 z_i^4 + \sum_{i,j}^* (b_{i,i} b_{j,j} + 2 b_{i,j}^2) z_i^2 z_j^2 + 2 \sum_{i,j}^* b_{i,i} b_{i,j} z_i^3 z_j \\
&\quad + \sum_{i,j,k}^* (2 b_{i,i} b_{j,k} + 4 b_{i,j} b_{i,k}) z_i^2 z_j z_k + \sum_{i,j,k,l}^* b_{i,j} b_{k,l} z_i z_j z_k z_l
\end{aligned}
$$

Firstly,

$$
\operatorname{Var}\left(\sum_{i=1}^m b_{i,i}^2 z_i^4 \right) = O(\sum_{i=1}^4 b_{i,i}^4)
$$

Secondly,

$$
\mathrm{E} \sum_{i,j}^* (b_{i,i} b_{j,j} + 2 b_{i,j}^2) z_i^2 z_j^2 = \sum_{i,j}^* (b_{i,i} b_{j,j} + 2 b_{i,j}^2) = (\sum_{i=1}^m b_{i,i})^2 - \sum_{i=1}^m b_{i,i}^2 + 2 \sum_{i,j=1}^n b_{i,j}^2 - 2 \sum_{i=1}^n b_{i,i}^2
$$

$$
\begin{aligned}
&(\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)z_i^2z_j^2)^2\\
=&\sum_{i,j}^{*}\sum_{k,l}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{k,k}b_{l,l}+2b_{k,l}^2)z_i^2z_j^2z_k^2z_l^2\\
=&\sum_{i,j}^{*}\sum_{k=i,l=j}^{*}+\sum_{i,j}^{*}\sum_{k=i,l\neq j}^{*}+\sum_{i,j}^{*}\sum_{k=j,l=i}^{*}+\sum_{i,j}^{*}\sum_{k=j,l\neq i}^{*}+\sum_{i,j}^{*}\sum_{k\notin\{i,j\},l=i}^{*}+\sum_{i,j}^{*}\sum_{k\notin\{i,j\},l=j}^{*}+\sum_{i,j}^{*}\sum_{k\notin\{i,j\},l\notin\{i,j\}}^{*}\\
=&2\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)^2z_i^4z_j^4+\sum_{i,j,l}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{i,i}b_{l,l}+2b_{i,l}^2)z_i^4z_j^2z_l^2\\
&+\sum_{i,j,l}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{j,j}b_{l,l}+2b_{j,l}^2)z_i^2z_j^4z_l^2\\
&+\sum_{i,j,k}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{k,k}b_{i,i}+2b_{k,i}^2)z_i^4z_j^2z_k^2\\
&+\sum_{i,j,k}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{k,k}b_{j,j}+2b_{k,j}^2)z_i^2z_j^4z_k^2\\
&+\sum_{i,j,k,l}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{k,k}b_{l,l}+2b_{k,l}^2)z_i^2z_j^2z_k^2z_l^2\\
=&2\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)^2z_i^4z_j^4+4\sum_{i,j,k}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{i,i}b_{k,k}+2b_{i,k}^2)z_i^4z_j^2z_k^2\\
&+\sum_{i,j,k,l}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{k,k}b_{l,l}+2b_{k,l}^2)z_i^2z_j^2z_k^2z_l^2
\end{aligned}
$$

Hence

$$
\begin{aligned}
&\mathrm{E}(\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)z_i^2z_j^2)^2\\
=&2\mu_4^2\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)^2+4\mu_4\sum_{i,j,k}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{i,i}b_{k,k}+2b_{i,k}^2)\\
&+\sum_{i,j,k,l}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{k,k}b_{l,l}+2b_{k,l}^2)\\
=&2(\mu_4^2-1)\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)^2+4(\mu_4-1)\sum_{i,j,k}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2)(b_{i,i}b_{k,k}+2b_{i,k}^2)\\
&+(\sum_{i,j}^{*}(b_{i,i}b_{j,j}+2b_{i,j}^2))^2
\end{aligned}
$$

Note that

$$\sum_{i,j}^{*}b_{i,i}^2b_{j,j}^2=\left(\sum_{i=1}^{m}b_{i,i}^2\right)^2-\sum_{i=1}^{m}b_{i,i}^4$$

$$
\begin{aligned}
\sum_{i,j,k}^{*} b_{i,i}^2 b_{j,j} b_{k,k} =& \sum_{i=1}^{m} \sum_{j\neq i} \sum_{k\notin\{i,j\}} b_{i,i}^2 b_{j,j} b_{k,k} \\
=& \sum_{i=1}^{m} \sum_{j\neq i} \sum_{k\neq i} b_{i,i}^2 b_{j,j} b_{k,k} - \sum_{i=1}^{m} \sum_{j\neq i} b_{i,i}^2 b_{j,j}^2 \\
=& 2\sum_{i=1}^{m} b_{i,i}^4 - (\sum_{i=1}^{m} b_{i,i}^2)^2 - 2(\sum_{i=1}^{m} b_{i,i}^3)(\sum_{i=1}^{m} b_{i,i}) + (\sum_{i=1}^{m} b_{i,i}^2)(\sum_{i=1}^{m} b_{i,i})^2
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \mathrm{E}(\sum_{i,j}^{*} b_{i,i} b_{j,j} z_i^2 z_j^2)^2 \\
=& 2\mu_4^2 \sum_{i,j}^{*} b_{i,i}^2 b_{j,j}^2 + 4\mu_4 \sum_{i,j,k}^{*} b_{i,i}^2 b_{j,j} b_{k,k} + \sum_{i,j,k,l}^{*} b_{i,i} b_{j,j} b_{k,k} b_{l,l} \\
=& 2(\mu_4^2 - 1) \sum_{i,j}^{*} b_{i,i}^2 b_{j,j}^2 + 4(\mu_4 - 1) \sum_{i,j,k}^{*} b_{i,i}^2 b_{j,j} b_{k,k} + (\sum_{i,j}^{*} b_{i,i} b_{j,j})^2 \\
=& 2(\mu_4^2 - 1) \left((\sum_{i=1}^{m} b_{i,i}^2)^2 - \sum_{i=1}^{m} b_{i,i}^4 \right) \\
& + 4(\mu_4 - 1) \left(2\sum_{i=1}^{m} b_{i,i}^4 - (\sum_{i=1}^{m} b_{i,i}^2)^2 - 2(\sum_{i=1}^{m} b_{i,i}^3)(\sum_{i=1}^{m} b_{i,i}) + (\sum_{i=1}^{m} b_{i,i}^2)(\sum_{i=1}^{m} b_{i,i})^2 \right) \\
& + (\sum_{i=1}^{m} b_{i,i}^2)^2 + (\sum_{i=1}^{m} b_{i,i})^4 - 2(\sum_{i=1}^{m} b_{i,i}^2)(\sum_{i=1}^{m} b_{i,i})^2 \\
=& (\sum_{i=1}^{m} b_{i,i})^4 + (2\mu_4^2 - 4\mu_4 + 3)(\sum_{i=1}^{m} b_{i,i}^2)^2 + (-2\mu_4^2 + 8\mu_4 - 6) \sum_{i=1}^{m} b_{i,i}^4 \\
& + (-8\mu_4 + 8)(\sum_{i=1}^{m} b_{i,i}^3)(\sum_{i=1}^{m} b_{i,i}) + (4\mu_4 - 6)(\sum_{i=1}^{m} b_{i,i}^2)(\sum_{i=1}^{m} b_{i,i})^2
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathrm{Var}(\sum_{i,j}^{*} b_{i,i}b_{j,j}z_i^2z_j^2)\\
=&(\sum_{i=1}^{m} b_{i,i})^4 + (2\mu_4^2 - 4\mu_4 + 3)(\sum_{i=1}^{m} b_{i,i}^2)^2 + (-2\mu_4^2 + 8\mu_4 - 6)\sum_{i=1}^{m} b_{i,i}^4\\
&+ (-8\mu_4 + 8)(\sum_{i=1}^{m} b_{i,i}^3)(\sum_{i=1}^{m} b_{i,i}) + (4\mu_4 - 6)(\sum_{i=1}^{m} b_{i,i}^2)(\sum_{i=1}^{m} b_{i,i})^2\\
&- (\sum_{i=1}^{m} b_{i,i})^4 - (\sum_{i=1}^{m} b_{i,i}^2)^2 + 2(\sum_{i=1}^{m} b_{i,i})^2(\sum_{i=1}^{m} b_{i,i}^2)\\
=&(2\mu_4^2 - 4\mu_4 + 2)(\sum_{i=1}^{m} b_{i,i}^2)^2 + (-2\mu_4^2 + 8\mu_4 - 6)\sum_{i=1}^{m} b_{i,i}^4\\
&+ (-8\mu_4 + 8)(\sum_{i=1}^{m} b_{i,i}^3)(\sum_{i=1}^{m} b_{i,i}) + (4\mu_4 - 4)(\sum_{i=1}^{m} b_{i,i}^2)(\sum_{i=1}^{m} b_{i,i})^2
\end{aligned}
$$

□

Appendices

Appendix A haha1

Proof of Proposition 1. We assume $0 < \alpha < 1$ since the case $\alpha = 0$ or 1 is trivial. Note that the condition implies $\int [\varphi(\mathbf{y}) - \alpha]\mathcal{N}_n(0, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) = 0$. Hence it suffices to prove $\varphi(\mathbf{y}) \geq \alpha$, a.s. We prove this by contradiction. Suppose $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha\}) > 0$. Then there exists a $\eta > 0$, such that $\lambda(\{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}) > 0$. We denote $E = \{\mathbf{y} : \varphi(\mathbf{y}) < \alpha - \eta\}$. From Lebesgue density theorem (Cohn, 2013, Corollary 6.2.6), there exists a point $z \in E$, such that, for each $\epsilon > 0$ there is a $\delta_\epsilon > 0$ such that

$$\left| \frac{\lambda(E^{\complement} \cap C_\epsilon)}{\lambda(C_\epsilon)} \right| < \epsilon,$$

where $C_\epsilon = \prod_{i=1}^{n} [z_i - \delta_\epsilon, z_i + \delta_\epsilon]$. We put

$$\epsilon = \left(\frac{\sqrt{\pi}}{\sqrt{2}\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3}.$$

Then for any $\phi > 0$,

$$
\begin{aligned}
\alpha \leq & \int_{\mathbb{R}^n} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \\
= & \int_{E \cap C_\epsilon} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) + \int_{E^\complement \cap C_\epsilon} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) + \int_{C_\epsilon^\complement} \varphi(\mathbf{y}) \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \\
\leq & \alpha - \eta + \int_{E^\complement \cap C_\epsilon} \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) + \int_{C_\epsilon^\complement} \mathcal{N}_n(z, \phi^{-1}\mathbf{I}_n)(d\mathbf{y}) \\
\leq & \alpha - \eta + \left(\frac{\phi}{2\pi} \right)^{n/2} \lambda(E^\complement \cap C_\epsilon) + 2n \left(1 - \Phi(\sqrt{\phi} \delta_\epsilon) \right) \\
\leq & \alpha - \eta + \left(\frac{\phi}{2\pi} \right)^{n/2} \epsilon (2\delta_\epsilon)^n + 2n \left(1 - \Phi(\sqrt{\phi} \delta_\epsilon) \right) \\
= & \alpha - \eta + \left(\frac{\sqrt{\phi} \delta_\epsilon}{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)} \right)^n \frac{\eta}{3} + 2n \left(1 - \Phi(\sqrt{\phi} \delta_\epsilon) \right).
\end{aligned}
$$

Putting

$$\phi = \left(\frac{\Phi^{-1}\left(1 - \frac{\eta}{6n}\right)}{\delta_\epsilon} \right)^2$$

yields the contradiction $\alpha \leq \alpha - (2/3)\eta$. This completes the proof.

$\square$

Proof of Proposition 2. For positive integer m, define $[m] = \{1, , \ldots, m\}$. For a set A, denote by $|A|$ its cardinality. We have

$$
\begin{aligned}
k_\kappa = & \left| \left\{ i \in [n-q] : \frac{\gamma_i^2}{\gamma_i + \kappa} - \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j \gamma_i}{\gamma_j + \kappa} > 0 \right\} \right| \\
= & \left| \left\{ i \in [n-q] : \frac{\gamma_i}{\gamma_i + \kappa} > \frac{1}{n-q} \sum_{j=1}^{n-q} \frac{\gamma_j}{\gamma_j + \kappa} \right\} \right|.
\end{aligned}
$$

Let X be a random variable uniformly distributed on $\{\gamma_1, \ldots, \gamma_{n-q}\}$. That is, $\Pr(X = \gamma_i) = 1/(n-q)$, $i = 1, \ldots, n-q$. Then it can be seen that

$$k_\kappa = (n-q) \Pr\left(\frac{X}{X+\kappa} > \mathrm{E}\left[\frac{X}{X+\kappa} \right] \right).$$

Hence we only need to verify

$$\Pr\left(\frac{X}{X+\kappa_1} > \mathrm{E}\left[\frac{X}{X+\kappa_1} \right] \right) \geq \Pr\left(\frac{X}{X+\kappa_2} > \mathrm{E}\left[\frac{X}{X+\kappa_2} \right] \right). \tag{2}$$

Let $Y = X/(X+\kappa_2)$. Then

$$\frac{X}{(X+\kappa_1)} = \frac{\kappa_2 Y}{\kappa_1 + (\kappa_2 - \kappa_1) Y} := f(Y).$$

Note that $f(Y)$ is increasing for $Y \geq 0$. Then the inequality (2) is equivalent to

$$\Pr\left(Y > f^{-1}\left(\mathrm{E}\, f(Y)\right)\right) \geq \Pr\left(Y > \mathrm{E}\, Y\right).$$

Hence we only need to verify $f^{-1}\left(\mathrm{E}\, f(Y)\right) \leq \mathrm{E}\, Y$, or equivalently, $\mathrm{E}\, f(Y) \leq f(\mathrm{E}\, Y)$. But the last inequality is a direct consequence of the concavity of $f(Y)$. This completes the proof.

□

Lemma 8. *Let $\mathbf{W}$ be an $N \times N$ positive semi-definite matrix. Let Z be an N dimensional sub-gaussian random vector with $\mathrm{E}\, Z = 0$, $\mathrm{Var}(Z) = \mathbf{I}_n$ and $\|Z\|_{\psi_2} \leq K$. For all $t > 0$,*

$$\Pr\left\{Z^\top \mathbf{W} Z >\right\} \leq e^{-t}.$$

Remark 1. This lemma is adapted from Hsu et al. (2012) Theorem 2.1 with minor modifications. Indeed, their result did not track the variance of Z.

Proof. Let $\mathbf{A} = \sqrt{\mathbf{W}}$. Let Z^* be a vector of N independent standard Gaussian random variables which are independent of Z.

$$\mathrm{E}[\exp(\lambda Z^{*\top} \mathbf{A} Z)] = \mathrm{E}\,\mathrm{E}[\exp(\lambda Z^{*\top} \mathbf{A} Z) | Z] = \mathrm{E}[\exp(\frac{\lambda^2}{2} Z^\top \mathbf{W} Z)]$$

□

Lemma 9.

Proof. □

Proof of Lemma 4.

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \mathrm{tr}(\mathbf{W}) \mathbf{I}_n\| \leq \|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \mathrm{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\| + \|\mathrm{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) - \mathrm{tr}(\mathbf{W}) \mathbf{I}_n\|$$

We have

$$\|\mathrm{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) - \mathrm{tr}(\mathbf{W}) \mathbf{I}_n\| = \max_{1 \leq i \leq n} |Z_i^\top \mathbf{W} Z_i - \mathrm{tr}(\mathbf{W})| \leq \sqrt{\sum_{i=1}^n \left(Z_i^\top \mathbf{W} Z_i - \mathrm{tr}(\mathbf{W})\right)^2}$$

From (Chen et al., 2010, Proposition A.1),

$$\mathrm{E}\left[\sum_{i=1}^n \left(Z_i^\top \mathbf{W} Z_i - \mathrm{tr}(\mathbf{W})\right)^2\right] = 2n\,\mathrm{tr}(\mathbf{W}^2) + \Delta n\,\mathrm{tr}(\mathbf{W} \circ \mathbf{W}) \leq (2 + \Delta) n\,\mathrm{tr}(\mathbf{W}^2).$$

Hence

$$\|\mathrm{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}) - \mathrm{tr}(\mathbf{W}) \mathbf{I}_n\| = O_P(\sqrt{n} \|\mathbf{W}\|_F).$$

Next we deal with

$$\|\mathbf{Z}^\top \mathbf{W} \mathbf{Z} - \mathrm{diag}(\mathbf{Z}^\top \mathbf{W} \mathbf{Z})\|$$

From (Vershynin, 2018, Lemma 5.2), there is a $1/4$-net $\mathcal{C}$ of the unit sphere S^{n-1} such that $|\mathcal{C}| \le 9^n$. By (Vershynin, 2018, Exercise 4.4.3),

$$\|\mathbf{Z}^\top \mathbf{W}\mathbf{Z} - \mathrm{diag}(\mathbf{Z}^\top \mathbf{W}\mathbf{Z})\| \le 2 \sup_{x\in\mathcal{C}} \left| x^\top \left(\mathbf{Z}^\top \mathbf{W}\mathbf{Z} - \mathrm{diag}(\mathbf{Z}^\top \mathbf{W}\mathbf{Z})\right) x \right|.$$

Fix $x \in \mathcal{C}$. Then

$$\left| x^\top \left(\mathbf{Z}^\top \mathbf{W}\mathbf{Z} - \mathrm{diag}(\mathbf{Z}^\top \mathbf{W}\mathbf{Z})\right) x \right| = \left| \sum_{i=1}^n \sum_{j\neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j \right|$$

Now we bound the moment generating function of $\sum_{i=1}^n \sum_{j\neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j$. We apply the decoupling technique in Vershynin (2018), Section 6.1. Let $\delta_1, \dots, \delta_n$ be independent Bernoulli random variables with $\Pr\{\delta_i = 0\} = \Pr\{\delta_i = 1\} = 1/2$. For any $\lambda \in \mathbb{R}$,

$$\begin{aligned}
\mathrm{E}\exp\left\{\lambda \sum_{i=1}^n \sum_{j\neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j\right\} =& \mathrm{E}\exp\left\{\mathrm{E}\left(4\lambda \sum_{i=1}^n \sum_{j=1}^n \delta_i(1-\delta_j) x_i x_j Z_i^\top \mathbf{W} Z_j \middle| \mathbf{Z}\right)\right\} \\
\le& \mathrm{E}\exp\left\{4\lambda \sum_{i=1}^n \sum_{j=1}^n \delta_i(1-\delta_j) x_i x_j Z_i^\top \mathbf{W} Z_j\right\} \\
=& \mathrm{E}\exp\left\{4\lambda (\sum_{i:\delta_i=1} x_i Z_i)^\top \mathbf{W} (\sum_{j:\delta_j=0} x_j Z_j)\right\} \\
\le& \max_{I\subset[n]} \mathrm{E}\exp\left\{4\lambda (\sum_{i\in I} x_i Z_i)^\top \mathbf{W} (\sum_{j\notin I} x_j Z_j)\right\},
\end{aligned}$$

where the first inequality follows from Jensen's inequality. Fix an $I \subset [n]$. From Vershynin (2018), Proposition 2.6.1, $\|\sum_{i\in I} x_i Z_i\|_{\psi_2} \le C_1 K$, $\|\sum_{j\notin I} x_j Z_j\|_{\psi_2} \le C_1 K$ for some absolute constant C_1. Then Vershynin (2018), Lemma 6.2.2 and Lemma 6.2.3 imply that there exist absolute constants C_2, C_3 such that,

$$\mathrm{E}\exp\left\{4\lambda (\sum_{i\in I} x_i Z_i)^\top \mathbf{W} (\sum_{j\notin I} x_j Z_j)\right\} \le \exp\left\{C_2 K^4 \|\mathbf{W}\|_F^2 \lambda^2\right\}$$

for all $|\lambda| \le C_3/(K^2\|\mathbf{W}\|)$. Note that this bound does not depend on $I \subset [n]$. It follows that

$$\mathrm{E}\exp\left\{\lambda \sum_{i=1}^n \sum_{j\neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j\right\} \le \exp\left\{C_2 K^4 \|\mathbf{W}\|_F^2 \lambda^2\right\},$$

for all $|\lambda| \le C_3/(K^2\|\mathbf{W}\|)$. Then applying Chernoff bound yields that, for any $t > 0$,

$$\begin{aligned}
\Pr\left(\left|\sum_{i=1}^n \sum_{j\neq i}^n x_i x_j Z_i^\top \mathbf{W} Z_j\right| > t\right) \le& \inf_{0<\lambda\le\frac{C_3}{K^2\|\mathbf{W}\|}} 2\exp\left\{-\lambda t + C_2 K^4 \|\mathbf{W}\|_F^2 \lambda^2\right\} \\
\le& 2\exp\left\{-\min\left(\frac{t^2}{4C_2K^4\|\mathbf{W}\|_F^2}, \frac{C_3 t}{2K^2\|\mathbf{W}\|}\right)\right\}.
\end{aligned}$$

This inequality, combined with union bound, yields

$$\begin{aligned}\Pr\Big(\|\mathbf{Z}^\top\mathbf{W}\mathbf{Z}-\operatorname{diag}(\mathbf{Z}^\top\mathbf{W}\mathbf{Z})\|>t\Big)\leq&\Pr\left(2\sup_{x\in\mathcal{C}}\left|x^\top\Big(\mathbf{Z}^\top\mathbf{W}\mathbf{Z}-\operatorname{diag}(\mathbf{Z}^\top\mathbf{W}\mathbf{Z})\Big)x\right|>t\right)\\ \leq&2\cdot 9^n\exp\left\{-\min\left(\frac{t^2}{16C_2K^4\|\mathbf{W}\|_F^2},\frac{C_3t}{4K^2\|\mathbf{W}\|}\right)\right\}.\end{aligned}$$

Thus, there exists a large $C>0$ such that for every $t>0$,

$$\Pr\Big(\|\mathbf{Z}^\top\mathbf{W}\mathbf{Z}-\operatorname{diag}(\mathbf{Z}^\top\mathbf{W}\mathbf{Z})\|>C(K^2(\sqrt{n}+t)\|\mathbf{W}\|_F+K^2(n+t^2)\|\mathbf{W}\|)\Big)\leq 2\exp\{-t^2\}.$$

Consequently, $\|\mathbf{Z}^\top\mathbf{W}\mathbf{Z}-\operatorname{diag}(\mathbf{Z}^\top\mathbf{W}\mathbf{Z})\|=O_P(K^2(\sqrt{n}\|\mathbf{W}\|_F+n\|\mathbf{W}\|))$. This completes the proof.

□

Appendix B haha2

Theorem 2. *Let $\zeta_1,\ldots,\zeta_d$ be iid random variables with mean 0 and variance 1, and assume $\mu_k:=\mathrm{E}(\zeta_1^k)$ is finite for $k\leq 8$. Let $\boldsymbol{\zeta}=(\zeta_1,\ldots,\zeta_d)^\top\in\mathbb{R}^d$. For $k=1,\ldots,K$, let $\mathbf{Q}_k=(q_{i_j}^{(k)})$ be a $d\times d$ symmetric matrix and let $\check{\mathbf{Q}}_k=\operatorname{diag}(q_{11}^{(k)},\ldots,q_{dd}^{(k)})$, $\hat{\mathbf{Q}}_k=\mathbf{I}_d-\check{\mathbf{Q}}_k$. Define $\hat{w}_k=\boldsymbol{\zeta}^\top\hat{\mathbf{Q}}_k\boldsymbol{\zeta}$, $\check{w}_k=\boldsymbol{\zeta}^\top\check{\mathbf{Q}}_k\boldsymbol{\zeta}-\operatorname{tr}(\mathbf{Q}_k)$, and*

$$W=\begin{pmatrix}\hat{w}_1\\ \check{w}_1\\ \vdots\\ \hat{w}_K\\ \check{w}_K\end{pmatrix}=\begin{pmatrix}\boldsymbol{\zeta}^\top\hat{\mathbf{Q}}_1\boldsymbol{\zeta}\\ \boldsymbol{\zeta}^\top\check{\mathbf{Q}}_1\boldsymbol{\zeta}-\operatorname{tr}(\mathbf{Q}_1)\\ \vdots\\ \boldsymbol{\zeta}^\top\hat{\mathbf{Q}}_1\boldsymbol{\zeta}\\ \boldsymbol{\zeta}^\top\check{\mathbf{Q}}_1\boldsymbol{\zeta}-\operatorname{tr}(\mathbf{Q}_1)\end{pmatrix}\in\mathbb{R}^{2K}.$$

Finally, let $Z\sim\mathcal{N}_{2K}(0,\mathbf{I}_{2K})$ and $\mathbf{V}=\operatorname{Cov}(W)$. There is an absolute constant $0<C<\infty$ such that

haha

Proof. Let $f:\mathbb{R}^{2K}\to\mathbb{R}$ be a four-times differentiable function. From xxx, there is a $4-times$ differentiable function $g:\mathbb{R}^{2K}\to\mathbb{R}$ satisfying the Stein identity

$$\mathrm{E}[f(W)]-\mathrm{E}[f(\mathbf{V}^{1/2}W)]=\mathrm{E}[\nabla^\top\mathbf{V}\nabla g(W)-W^\top\nabla g(W)]$$

and

$$\left|\frac{\partial^k g(\mathbf{x})}{\prod_{j=1}^k\partial x_{i_j}}\right|\leq\frac{1}{k}\left|\frac{\partial^k f(\mathbf{x})}{\prod_{j=1}^k\partial x_{i_j}}\right|\quad\text{for all }\mathbf{x}=(x_1,\ldots,x_{2K})^\top\in\mathbb{R}^{2K},\ k=1,2,3,\text{ and }i_j\in\{1,\ldots,2K\}.$$

To prove the theorem, we bound

$$S=\mathrm{E}[\nabla^\top\mathbf{V}\nabla g(W)-W^\top\nabla g(W)].$$

Next, we use exchangeability. Let $\boldsymbol{\zeta}' = (\zeta_1', \ldots, \zeta_d')^\top$ be an independent copy of ζ, and let $\underline{i} \in \{1, \ldots, d\}$ be an independent and uniformly distributed random index. Define the vector $W' \in \mathbb{R}^{2K}$ exactly as we defined W, except that $\boldsymbol{\zeta}_{\underline{i}}$ is replaced with $\boldsymbol{\zeta}_{\underline{i}}'$ throughout. More precisely, let $e_i \in \mathbb{R}^d$ be the ith standard basis vector in $\mathcal{R}^d$ and define

$$\begin{aligned}\hat{w}_k' =&(\boldsymbol{\zeta} + (\zeta_{\underline{i}}' - \zeta_{\underline{i}})e_{\underline{i}})^\top \hat{\mathbf{Q}}_k(\boldsymbol{\zeta} + (\zeta_{\underline{i}}' - \zeta_{\underline{i}})e_{\underline{i}})\\ =&\hat{w}_k + 2(\zeta_{\underline{i}}' - \zeta_{\underline{i}})e_{\underline{i}}^\top \hat{\mathbf{Q}}_k \boldsymbol{\zeta},\end{aligned}$$

$$\begin{aligned}\check{w}_k' =&(\boldsymbol{\zeta} + (\zeta_{\underline{i}}' - \zeta_{\underline{i}})e_{\underline{i}})^\top \check{\mathbf{Q}}_k(\boldsymbol{\zeta} + (\zeta_{\underline{i}}' - \zeta_{\underline{i}})e_{\underline{i}}) - \mathrm{tr}(\mathbf{Q}_k)\\ =&\check{w}_k + e_{\underline{i}}^\top \check{\mathbf{Q}}_k e_{\underline{i}}((\zeta_{\underline{i}}')^2 - \zeta_{\underline{i}}^2),\end{aligned}$$

for $k = 1, \ldots, K$. Then $W' = (\hat{w}_1', \check{w}_1', \ldots, \hat{w}_K', \check{w}_K')^\top \in \mathbb{R}^{2K}$. Its straightforward to verify that

$$\mathrm{E}(\hat{w}_k' - \hat{w}_k|\boldsymbol{\zeta}) = -\frac{2}{d}\hat{w}_k, \quad \mathrm{E}(\check{w}_k' - \check{w}_k|\boldsymbol{\zeta}) = -\frac{1}{d}\check{w}_k.$$

Then

$$\mathrm{E}(W' - W|\boldsymbol{\zeta}) = -\Lambda_K W,$$

where

$$\Lambda_1 = \begin{pmatrix} \frac{2}{d} & 0 \\ 0 & \frac{1}{d} \end{pmatrix}, \quad \Lambda_K = \begin{pmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Lambda_1 \end{pmatrix} \in \mathbb{R}^{2K \times 2K}.$$

By exchangeability, we have

$$\begin{aligned}0 =&\frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') + \nabla g(W))]\\ =&\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top} \nabla g(W)] + \frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') - \nabla g(W))]\\ =&-\mathrm{E}[W^\top \nabla g(W)] + \frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') - \nabla g(W))].\end{aligned}$$

That is,

$$\mathrm{E}[W^\top \nabla g(W)] = \frac{1}{2}\,\mathrm{E}[(W' - W)^\top \Lambda_K^{-\top}(\nabla g(W') - \nabla g(W))].$$

Apply Taylor's theorem,

$$
\begin{aligned}
&W^\top \nabla g(W) \\
=&\frac{1}{2}\sum_{i,j=1}^{2K} \Lambda_{K,ii}^{-1} D^{ij} g(W)(w_i' - w_i)(w_j' - w_j) + \frac{1}{4}\sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W)(w_i' - w_i)(w_j' - w_j)(w_k' - w_k) \\
&+ \frac{1}{12}\sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W)(w_i' - w_i)(w_j' - w_j)(w_k' - w_k)(w_l' - w_l) \\
=&\frac{1}{2}\operatorname{tr}[(W' - W)(W' - W)^\top \Lambda_K^{-\top} \nabla^2 g(W)] + \frac{1}{4}\sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W)(w_i' - w_i)(w_j' - w_j)(w_k' - w_k) \\
&+ \frac{1}{12}\sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W)(w_i' - w_i)(w_j' - w_j)(w_k' - w_k)(w_l' - w_l),
\end{aligned}
\tag{3}
$$

where $t^* \in [0,1]$. Also by exchangeability,

$$
\mathrm{E}[(W' - W)(W' - W)^\top] = 2\,\mathrm{E}[W(W - W')^\top] = 2\,\mathrm{E}[WW^\top \Lambda_K^\top] = 2\mathbf{V}\Lambda_K^\top.
$$

It follows that

$$
\mathrm{E}[\nabla^\top \mathbf{V} \nabla g(W)] = \mathrm{E}\operatorname{tr}[\mathbf{V}\nabla^2 g(W)] = \frac{1}{2}\mathrm{E}\operatorname{tr}[\mathrm{E}[(W' - W)(W' - W)^\top]\Lambda_K^{-\top}\nabla^2 g(W)]
$$

Thus,

$$
\begin{aligned}
S =& \mathrm{E}[\nabla^\top \mathbf{V}\nabla g(W) - W^\top \nabla g(W)] \\
=&\frac{1}{2}\mathrm{E}\operatorname{tr}[\mathrm{E}[(W' - W)(W' - W)^\top]\Lambda_K^{-\top}\nabla^2 g(W)] - \frac{1}{2}\mathrm{E}\operatorname{tr}[(W' - W)(W' - W)^\top \Lambda_K^{-\top}\nabla^2 g(W)] \\
&- \frac{1}{4}\mathrm{E}\sum_{i,j,k=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijk} g(W)(w_i' - w_i)(w_j' - w_j)(w_k' - w_k) \\
&- \frac{1}{12}\mathrm{E}\sum_{i,j,k,l=1}^{2K} \Lambda_{K,ii}^{-1} D^{ijkl} g(t^*(W' - W) + W)(w_i' - w_i)(w_j' - w_j)(w_k' - w_k)(w_l' - w_l).
\end{aligned}
$$

□

References

Chen, S. X., Zhang, L., and Zhong, P. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 105(490):810–819.

Cohn, D. L. (2013). *Measure Theory*. Birkhauser Advanced Texts Basler Lehrbucher. Birkhuser Basel, 2 edition.

Goeman, J. J., van de Geer, S. A., and van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):477–493.

Hsu, D., Kakade, S., and Zhang, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(0).

Jiang, J. (1996). Reml estimation: asymptotic behavior and related topics. *Annals of Statistics*, 24(1):255–286.

Kass, R. E. and Wasserman, L. (1995). A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Maruyama, Y. and George, E. I. (2011). Fully bayes factors with a generalized g -prior. *Ann. Statist.*, 39(5):2740–2765.

Shang, Z. and Clayton, M. K. (2011). Consistency of bayesian linear model selection with a growing number of parameters. *Journal of Statistical Planning and Inference*, 141(11):3463–3474.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.